

TITLE: **GRADING CONTACT LENS COMPLICATIONS: THE EFFECT OF KNOWLEDGE ON GRADING ACCURACY**

(Running head title): Effect of Knowledge on grading accuracy

Authors: **Genís Cardona, PhD***
Carme Serés, MSc, FIACLE*

*Research Group of Centre Universitari de la Visió
Optics and Optometry Department, Technical University of Catalonia

Corresponding Author: Genís Cardona
Escola Universitària d'Òptica i Optometria de Terrassa
Violinista Vellsolà, 37
08222 Terrassa, Catalonia, Spain

E-mail address: gcardona@oo.upc.edu

ABSTRACT

Purpose: Grading contact lens complications is a time-effective process which, if performed accurately, may improve patient management significantly. Multiple factors have been identified that influence grading accuracy, such as the knowledge, training and experience of the observer. The aim of this study was to further explore the effect of knowledge on grading accuracy, both in terms of intensity and specificity, and to do so by avoiding the need to require the subjects to interpolate between whole number grades. *Material and Methods:* Optometry students were divided into three knowledge groups according to their academic progress and in such a way that knowledge intensity and specificity could be evaluated separately. A vertical visual analogue scale was devised to allow subjects to grade three different conditions of medium severity (position 50 of the scale) from the Institute for Eye Research grading scale, mainly, bulbar hyperaemia, limbal vascularisation and lid redness. *Results:* Bulbar hyperaemia was graded lowest (mean = 21.5; SD = 1.4), followed by limbal vascularisation (mean = 40.7; SD = 1.6) and lid redness (mean = 51.2; SD = 1.3) and these differences were found to be statistically significant (ANOVA $p = 0.000$). Knowledge was found to influence the grading process, resulting in statistically significant differences in the overall grading performance of the three knowledge groups (ANOVA, $p = 0.048$). Furthermore, even if knowledge intensity failed to improve grading accuracy over a certain threshold, a combination of knowledge intensity and specificity resulted in an enhancement of the homogeneity of the grading data. *Conclusions:* Knowledge intensity and specificity both contribute to improve grading skills, albeit through different mechanisms. An intermediate knowledge of contact lens complications is required to attain good grading accuracy and a basic training in pathology is also advised.

KEY WORDS

Contact lens complications; Grading scales; Grading accuracy; Interpolation skills; Knowledge

I. INTRODUCTION

The grading of contact lens complications is a complex and essential process that stems from the need to correctly identify the severity of a condition and, therefore, to allow for the most appropriate form of treatment. The prognosis of many contact lens complications depends on the ability of the clinician to decide upon the actual severity and urgency of the condition and to act accordingly, either through treatment or referral.

Many attempts have been made to objectivise the grading process. Image analysis software permits the direct evaluation of a real image through a set of established parameters which, if properly adjusted, may provide useful information about the severity of the particular condition for which they have been defined. Most of these techniques, however, are constricted to a research lab and to a few, very specific, conditions, mainly bulbar or conjunctival redness, corneal oedema and palpebral roughness. To cite a few, O'Donnell and Wolffsohn¹ digitalized the corneal oedema images from the Institute for Eye Research (formerly CCLRU) grading scales and compared the intensity of grey of the real images with that of their scale, thus obtaining a value of severity. These authors disclosed that the severity of corneal oedema presented a better correlation to the anteroposterior gradient than to the overall intensity or greyness of the image. Papas² investigated the level of agreement between the subjective grades of conjunctival hyperaemia and a set of variables supplied by the image analysis software, encountering a moderate correlation when the level of redness alone was taken into account, but a much better agreement between subjective and objective grades when the degree of tortuosity of the blood vessels, their diameter, and the size of the area which was covered by them was evaluated. These results led to the utilization of edge detection software³ and fractal analysis⁴ to precisely determine the profile of conjunctival blood vessels. Millán *et al*⁵, also basing their research on the Institute of Eye Research (IER) grading scales, firstly homogenized the size of the area under study by the application of a

complex polynomial transformation to the images, and, later, translated the RGB (red, green and blue) scale into HSI (hue, saturation and intensity) values, thus allowing further mathematical analysis of each condition. Sorbara and colleagues⁶, as a first attempt to approach research and clinical practice, addressed the need to obtain immediate objective grades of conjunctival hyperaemia by means of a photometer mounted on a slit lamp.

Notwithstanding these and other examples, however, grading of contact lens complications is still a predominantly subjective process which involves the comparison of a real live image with a set of previously selected images, representing various degrees of severity, a process which may be further modulated by the experience, training and knowledge of the observer.

The most commonly used contact lens complications grading scales are the real image IER grading scale and the artistic Efron scale. The IER grading scale represents 10 different anterior eye complications in four discrete and progressive degrees of severity ranging from “very slight” (step 1) to “severe” (step 4)⁷. It might be of interest to note that the IER scale lacks a step 0, that is, does not display an image of the healthy or “normal” appearance of each particular condition or ocular structure under study. The Efron scale is artistic in nature, and offers two different versions of the scale⁸. The first version consists of drawings of 16 different conditions in five severity steps, ranging from “normal” (step 0) to “severe” (step 4). The second version presents the same 16 ocular complications in a computer assisted *morph* display whereupon each condition moves from step 0 to step 4 in an apparently continuous way by morphing each 0.1 increment with the immediately anterior and posterior neighbours.

Both the IER scale and the Efron scale have their supporters and detractors. The IER scale has been criticized for the lack of perfect homogeneity between images representing the same condition, either in terms of different illumination conditions or variable size of the area under display⁹. The Efron scale overcomes these difficulties by encouraging artistic clarity and licence so as to emphasize and isolate the condition that is being evaluated. This might be considered

a departure from the real life situation in as much as different conditions feed and depend on each other and, therefore, occur simultaneously and should appear as such in a single image.

The Efron scale has been employed to test the effect of knowledge, experience and training of the observer on grading accuracy and reliability¹⁰. Knowledge was defined as extensive education in optometry, ophthalmology, contact lenses and contact lens associated complications; experience referred to the previous utilization of Efron grading scales; training required the observer to learn how to use the grading scale by assisting to one or more sessions of instruction. By comparing the grading skills of a group of optometrists with those of a group of non-optometrists, Efron and co-workers concluded that non-optometrists could grade with as much accuracy as optometrists, but their grades were less reliable (understood as similarity of grades between two consecutive sessions), mainly when they were presented with very mild conditions. Additionally, non-optometrists had difficulty in interpolating their grades into 0.1 increments, preferring unit or half-unit steps. Regarding training, a parallel study revealed a slight improvement in accuracy when optometrists were trained in between consecutive sessions, but the results failed to achieve statistical significance¹¹.

The main handicap of the grading process may be encountered in its very nature: the discreet quantification of a continuous, progressive condition. Bailey and colleagues¹² addressed the need to correctly determine the amplitude of the steps of a grading scale, as well as to ensure that this amplitude remained unaltered throughout the scale. A scale with few, widely spaced intervals will offer good intersubject and intrasubject repeatability but will display low sensibility to reflect slight changes in the progression of the condition. The opposite situation results in a finely tuned scale with a poor repeatability in the allocation of grades. These authors also acknowledged a different behaviour when observers were asked to grade conditions whose severity was positioned towards either end of the scale than when it fell somewhere towards the middle of the scale. Finally, Bailey *et al* recommended grading to the

nearest 0.1 increment in order to increase the precision of the grading process, even when the scale was divided in whole number steps.

The aim of this study was to further explore the mechanisms governing grading accuracy of contact lens complications. The knowledge of the observer was monitored and its influence on the grading process was investigated. Thus, three different groups of knowledge were defined in order to evaluate whether it was knowledge intensity or knowledge specificity the leading factor which influenced grading skills. Besides, as different levels of knowledge were found to result in either keenness or reluctance to interpolate between whole number grades, interpolation might be considered an intermediate variable when assessing grading accuracy. Therefore, a continuous analogue scale was developed to facilitate statistical analysis and to provide a means to negate the effect of interpolation as subjects were only in possession of end of scale clues for each condition to be graded.

II. METHODS

Subjects

Optometry students from the Technical University of Catalonia were recruited for this study, after briefly explaining the nature of the experiment to them. A total of 60 students participated in the study, with a mean age of 21.23 years (ranging from 20 to 25). 42 of the subjects were female and 18 were male.

Subjects were allocated into three different groups. Group 1 comprised 16 students enrolled to the Basic Contact Lens course. These students had a basic knowledge of the fundamental properties of contact lens materials, of normal corneal anatomy and physiology and of anterior ocular biometry. 22 students were assigned to Group 2. These students had an advanced knowledge regarding both RGP and soft contact lens fitting procedures and were able to correctly identify the most common contact lens complications. Finally, Group 3 included 22 last year students who, apart from an extensive theoretical and practical knowledge on the majority of contact lens topics, also had followed a course on Ocular Pathology which gave them a better understanding of the aetiology, diagnosis, treatment and prognosis of many commonly encountered pathologies of the anterior eye, including the three conditions under evaluation. It was assumed that the level of knowledge in the general field of contact lenses increased in similar steps from Group 1 to Group 3, thus defining the term “knowledge intensity”. Group 3, as well as increasing their general contact lens related knowledge, also gained an additional and specific knowledge on ocular pathology which gave rise to the term “knowledge specificity”.

Although some Group 3 students tended to be slightly older than their counterparts from Groups 1 and 2, no statistically significant difference between the ages of the 3 groups could be ascertained. None of the subjects had had any previous experience or training in the use of

grading scales but students from Groups 2 and 3 were aware of their existence. The Declaration of Helsinki tenets were followed throughout the study.

Grading Scales

Three conditions from the IER grading scale were selected for this study: bulbar hyperaemia, limbal vascularisation and lid redness. Subjects used a vertical visual analogue scale (VVAS) to grade each of the three conditions. This scale consisted on a vertical straight line without any markings anywhere on its 100 millimetres of length. The top of the scale was labelled as “very severe” and the bottom as “normal/healthy” (see **Figure 1**).

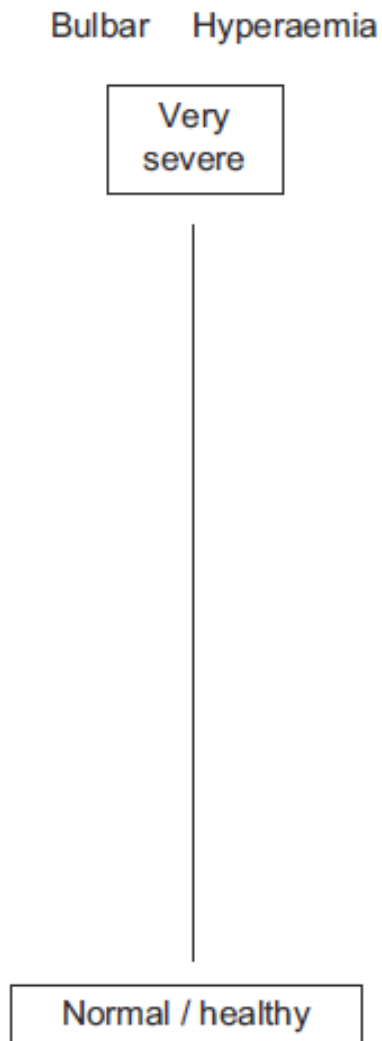


FIGURE 1 Example of a vertical visual analog scale (VVAS) used for the evaluation of bulbar hyperemia.

The image that had to be graded was displayed on the right hand side of a 20 inch computer screen set to a resolution of 1280 per 1024 pixels and a 32 bit colour configuration. The size of this image corresponded to 624 per 465 pixels. For each of the three conditions, the image to be graded was selected to represent step 2 of the IER scale, labelled as “slight” on the printed version of the scale. This image, as will be seen below, would theoretically be the equivalent to a mark at the exact centre of the VVAS (50 millimetres).

In order to assist in the grading process and to offer some visual clue to the observers, two additional images were presented on the left hand side of the computer screen, each subtending a resolution of 540 per 407 pixels. The superior image corresponded to a step 4 of the IER scale, therefore implying a mark of 100 on the VVAS. The inferior image portrayed a normal or healthy eye, that is, it was equivalent to a mark of 0 on the VVAS or a hypothetical step 0 on the IER scale (see **Figure 2 a, b, c**).



FIGURE 2 (A) Grading bulbar hyperemia. Image under evaluation (right) and end of scale clues (left). (B) Grading limbal vascularization. Image under evaluation (right) and end of scale clues (left). (C) Grading lid redness. Image under evaluation (right) and end of scale clues (left).

As noted above, the IER Grading Scale does not possess a step 0, instead choosing to start at the step 1 or very slight degree of severity state. Therefore, the image depicting a normal or healthy eye had to be artificially derived from the image which the IER scale allocated to step 1. This was accomplished through a computer assisted manipulation of the images (Adobe® Photoshop® CS3 Extended, version 10.0). On the images which illustrated bulbar hyperaemia and limbal vascularisation this manipulation consisted on the removal of any significant blood vessel from the anterior ocular surface and a general whitening of the exposed sclera. Lid redness was subsequently altered by decreasing the levels of red from the image and by manually erasing the most prominent of the blood vessels, thus uniformizing the overall colour of the displayed area. This process converted the original 4 step IER grading scale into a 5 step scale. Step 2 and step 4 images remained unaltered.

Procedure

Each subject was asked to grade the three conditions in a random order and only once. Subjects were allowed 30 seconds to grade each condition, during which they could freely compare the image under evaluation with the images indicative of a very severe condition and of a normal or healthy eye. Grades were assigned by marking the desired location on the VVAS. Room illumination and other conditions were constant for all subjects and throughout the experiment. All subjects were asked to sit at approximately 50 centimetres from the computer screen and were guided through the procedure by an experimenter who was unaware of the group number to which each one of the subjects was allocated. Grading marks (ranging from 0 to 100) were later measured from each VVAS by a research assistant who was ignorant of the purpose of the study.

Data Analysis

Efron and co-workers¹³ justified the use of parametric statistical analysis on grading data on the assumption of the normality of its distribution. Besides, it is acknowledged that multiple analysis of variance (ANOVA) tests tend to be robust enough to allow for small departures from a perfect Gaussian shape¹⁴. Therefore, the effect of the factors “Condition” and “Knowledge Group” on the dependent variable “Grading Score” was evaluated with a multifactor ANOVA. A Fisher’s Least Significant Difference (LSD) procedure was also utilized to explore statistical significance when paired groups of data were compared. A *p-value* of 0.05 or less was considered to denote statistical significance throughout the study.

III. RESULTS

Pooling together the grading data from the three groups on each condition (see **Figure 3 a**), it was disclosed that bulbar hyperaemia was graded lowest (mean = 21.5; SD = 1.4), followed by limbal vascularisation (mean = 40.7; SD = 1.6) and lid redness (mean = 51.2; SD = 1.3). An ANOVA test revealed these differences to be statistically significant ($p = 0.000$). Knowledge was found to influence the grading process, resulting in statistically significant differences in the overall grading performance of the three knowledge groups (ANOVA, $p = 0.048$). As this level of significance approaches the 0.05 threshold, groups were evaluated in pairs, using a Fisher's LSD test. Consequently, statistically significant differences were revealed between Group 1 and 2 and between Group 1 and 3 but Groups 2 and 3 were found to be similar. Indeed, Group 1 tended to assign the lowest grades (mean = 34.7; SD = 1.6), followed by Group 3 (mean = 39.0; SD = 1.4) and Group 2 (mean = 39.7; SD = 1.4) (see **Figure 3 b**).

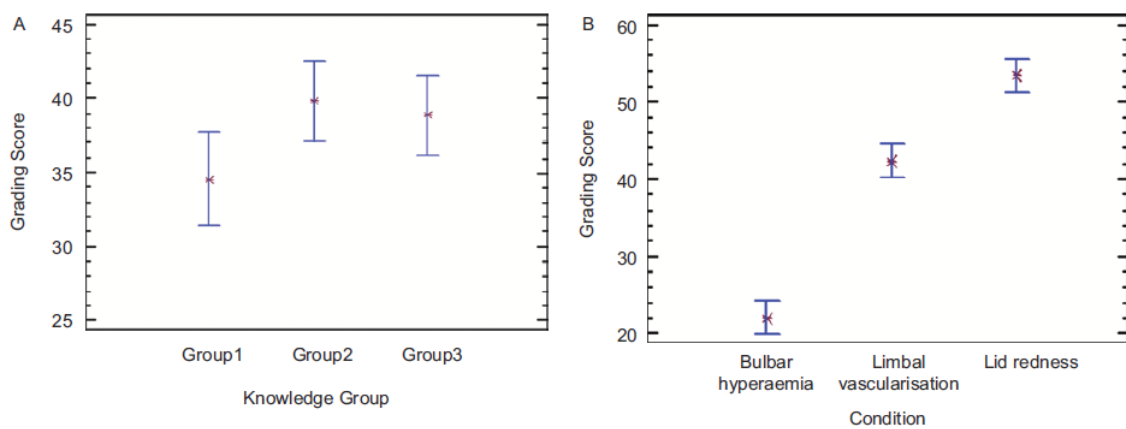


FIGURE 3 (A) Overall grading performance for each knowledge group (means and 95% LSD intervals). (B) Overall grading scores for each condition (means and 95% LSD intervals).

Although grading scores were rather different, grading behaviour was found to exhibit a similar pattern for limbal vascularisation and lid redness, that is, in both conditions grading scores from Group 1 tended to decrease when compared with those from Groups 2 and 3, whereas bulbar hyperaemia followed a different pattern, in which grading scores from Group 1 increased with regards to those from Groups 2 and 3 (see **Figure 4 a**).

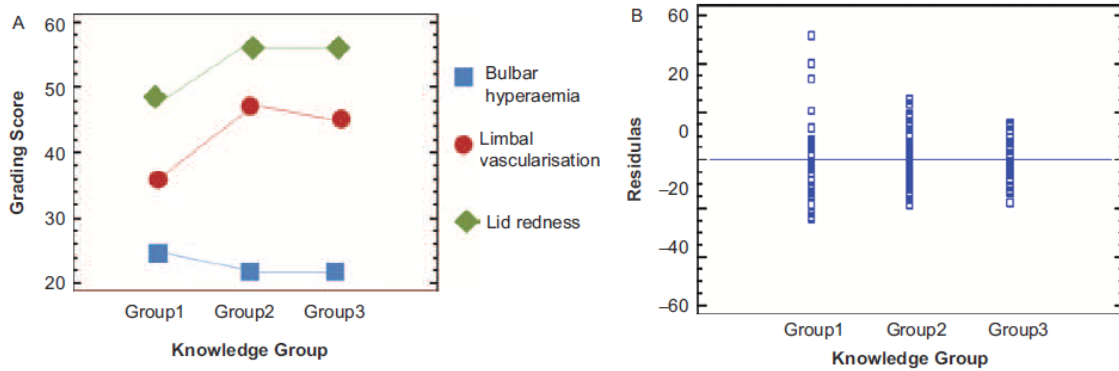


FIGURE 4 (A) Interaction plot for grading scores and condition, depicting grading behavior. (B) Residual plot of the grading scores for each knowledge group.

The analysis of the residual plot for each knowledge group (see **Figure 4 b**) exposed that, although grading scores from Groups 2 and 3 were comparable, results from Group 3 were more homogeneous, that is, discrepancies between subjects were smaller than those from Group 2. Grading scores from Group 1 presented the largest degree of heterogeneity. In addition, Group 1 comprised a few subjects whose grading behaviour was extreme enough to clearly break the cohesion of the group as a whole.

IV. DISCUSSION

Accuracy is defined by International standard ISO 5725-1 (1994) as “the closeness of agreement between a test result and the accepted reference value”¹⁵. Accuracy in the grading process is of paramount importance for the appropriate diagnosis and management of ocular conditions which, even if derived from contact lens use (or abuse), often present themselves as a subtle change in colour or vascular tortuosity that might prove difficult to discriminate from a baseline or healthy appearance.

The influence of knowledge in the process of grading contact lens complications has been previously explored. Efron and co-workers¹⁰ enlisted optometrists and non-optometrists for their study, and discovered that, even if non-optometrists could grade with as much accuracy as optometrists, they were more reluctant to do so by interpolating to the nearest 0.1 step of the scale. It could be speculated whether, had this limitation not existed, non-optometrists would have graded more accurately than optometrists, at least when the condition to be graded fell in between the unit or half-unit positions. It follows that knowledge, as well as interpolation skills, could be assumed to influence grading accuracy. Both accuracy and interpolation skills would be dependent variables of knowledge but, at the same time, it is possible that accuracy relies on the interpolation skills of the subject whose grading proficiency is being evaluated. If this was the case, interpolation skills could be considered to behave like an intermediate variable and to compromise the external validity of the conclusions.

Although the veracity of these assumptions calls for further evaluation, we felt the need to develop a method whereupon the influence of interpolation skills could be controlled or even negated. A vertical visual analogue scale was implemented to allow subjects to grade anywhere they considered appropriate between the bottom (labeled as “normal/healthy”) and the top of the scale (labeled as “very severe”). This scale consisted of a straight 100 millimeter line with no marks denoting steps. A vertical scale was preferred over a horizontal one to

prevent the reported hand dominance effect whereupon subjects tend to allocate more grades to either the right or the left hand side of a horizontal scale depending on their hand dominance¹⁶. The use of a continuous grading technique, as opposed to the discrete 4 or 5 steps grading scales, also offered the advantage of permitting the choice of a more robust statistical approach for the analysis of the results.

Even if it could be acknowledged that the artificial manipulation of an established and standardized scale such as the IER grading scale is not an optimal procedure, it was considered to be the lesser evil solution when attempting to recreate step 0 or normal/healthy eye conditions. The alternative, which consisted on utilizing an independent ocular image to illustrate the bottom end of the scale, was thought to incorporate unpredictable factors into the study, such as differences in the observable area, magnification, illumination and others, which could potentially confound the appearance of the condition or the judgement of the observers. Common sense was applied in order to transform step 1 images into step 0 images, and diverse photographs depicting normal and healthy eyes were perused. The result of the transformation was deemed to be satisfactory when, upon consultation, several independent external optometrists reported each one of the images to represent a healthy eye. Although the Efron scale incorporates a grade 0, thus saving the need for further photographic manipulation, the real life nature of the IER scale was preferred over the artistic licence of the Efron scale in order to present the subjects with images similar to those they had encountered throughout their studies.

The effect of knowledge was investigated in terms of its intensity and specificity. While Groups 1 and 2 only differed in knowledge intensity, both knowledge intensity and specificity was gained in Group 3.

The overall effect of knowledge on grading skills was more noticeable between Groups 1 and 2 than between Groups 2 and 3, where no statistically significant difference could be

established. It could be speculated that knowledge intensity improves grading skills, but only until a certain *plateau* has been reached, after which an equivalent increase in intensity (such as progressing from Group 2 to Group 3) did not lead to any visible improvement in the ability of students to grade with better accuracy.

It is interesting to note that an increase in knowledge intensity seems to positively affect the homogeneity of the grading scores within a given group (homogeneity relates to the assumption that the properties of any one part of an overall dataset are the same as any other part), an effect that is even discernible when comparing Groups 2 and 3, otherwise similar in terms of grading accuracy. This behaviour could be attributed to the homogenizing influence of a common academic curriculum in which all subjects within a group received exactly the same information regarding contact lens fitting and complications as well as, in the case of Group 3, basic pathology notions. Even if knowledge intensity failed to improve grading skills above a certain level, knowledge specificity was the most probable factor governing the improvement in the homogeneity of the grading scores which was obtained in Group 3.

Even though the results pointed towards a common grading performance for each particular group, it is interesting to note that, within each group, the grading proficiency of many of the subjects, when explored individually, was found to be consistent for all three conditions. That is, not only some observers tended to grade all three conditions with more accuracy than others but also, in doing so, they allocated either only higher or only lower grades throughout the grading session. This behaviour, which is in agreement with the findings of Efron *et al*¹³, was detected in a similar proportion in all three study groups and, therefore, could not be accounted for as a possible bias of knowledge intensity or specificity. It could perhaps be relevant to examine the mechanisms governing grading consistency and to investigate whether, with independence of their knowledge, consistent graders, as opposed to disperse

ones, would obtain some benefit from one or various training sessions devoted to hone their natural grading skills.

All groups tended to bestow bulbar hyperaemia with a grade which was significantly lower than the intended grade of 50 that the image was considered to represent (mean grade for bulbar hyperaemia = 21.52). The other conditions received grades approaching the mark of 50, with the possible exception of limbal vascularisation for Group 1 (mean = 34.19).

The particular case of bulbar hyperaemia might be explained in terms of the introduction, in the digital treatment of the image depicting a healthy eye, of some unknown artefact which could interfere with the appraisal of the subjects. Indeed, although not only the procedure followed in the conversion of step 1 images into step 0 images was common for all three conditions, but also the resulting images were deemed suitable by a team of independent optometrists, the total absence of blood vessels and the unnatural white appearance of the conjunctival area might have led the subjects to underestimate the severity of the image under evaluation. In addition, the possibility that not all step 2 images from the IER grading scale actually represent conditions at the exact 50 mark of severity must be accepted⁹.

Grading accuracy greatly depends on the complexity of the condition to be graded, especially if time is limited. Efron and McCubbin¹⁷ disclosed a significantly greater grading precision when observation times increased from 0.1 s to 2 s, but little further improvement when subjects were allowed 60 s to grade a single condition. However, the same authors acknowledged that grading pathologically complex conditions such as blepharitis or meibomian gland dysfunction would benefit of longer grading times. Although lid redness might be considered to be slightly more complex to grade than bulbar hyperaemia or limbal vascularisation, allowing for the fact that lid roughness as well as actual redness may be assessed, a grading time of 30 s was judged to be sufficient to prevent time to be a factor in the analysis of the results.

Grading behaviour is also known to be influenced by the order in which conditions are presented¹³. The appraisal of the severity of a particular condition may be influenced, either way, by the severity of the condition that was graded immediately before it. However, as images were displayed in a random order, this effect was avoided in the present study.

Therefore, the actual reasons for the peculiar grading results for bulbar hyperaemia remain unclear and demand further investigation. However, it may be assumed that these findings are independent of both knowledge intensity and specificity, as they emerged in all three knowledge groups, and are believed not to compromise the validity of the study.

In conclusion, both knowledge intensity and specificity seem to positively influence grading accuracy for any given subject, as well as the homogeneity of the grading results of each knowledge group as a whole. An intermediate knowledge in contact lens related complications is required to improve grading accuracy and, although a proficiency in ocular pathology may not be essential, it is nonetheless advisable. Grading contact lens complications is a fast, easy, cost-effective and efficient process that greatly improves the diagnosis, treatment and, therefore, outcome of these complications. Clinicians are encouraged to grade the conditions they observe and to do so with as much accuracy as possible.

V. ACKNOWLEDMENTS

The authors would like to acknowledge the kind assistance of Miquel Ralló, from the Department of Applied Mathematics of the Technical University of Catalonia, in the statistical analysis of the data

VI. REFERENCES

1. O'Donnell C, Wolffsohn JS. Grading of corneal transparency. *CLAE*. 2004; 27:161-70.
2. Papas EB. Key factors in the subjective and objective assessment of conjunctival erythema. *Invest Ophthalmol Vis Sci*. 2000; 41:687-691.
3. Wolffsohn JS, Purslow C. Clinical monitoring of ocular physiology using digital analysis. *CLAE*. 2003; 26:27-35.
4. Schulze MM, Hutchings N, Simpson TL. The use of fractal analysis and photometry to estimate the accuracy of bulbar redness grading scales. *Invest Ophthalmol Vis Sci*. 2008; 49:1398-1406.
5. Millán MS, Pérez-Cabré E, Abril HC, Valencia E. Evaluación objetiva de la hiperemia de la conjuntiva tarsal superior mediante análisis de imagen. Ensayo preliminar. *Rev Esp Contact*. 2005; 12:9-15.
6. Sorbara L, Simpson T, Duench S, Schulze M, Fonn D. Comparison of an objective method of measuring bulbar redness to the use of traditional grading scales. *CLAE*. 2007; 30:53-59.
7. Institute for Eye Research. Appendix B. In: Phillips AJ, Speedwell L, eds. *Contact Lenses*, 5th ed. Edinburgh, UK: Butterworth-Heinemann; 2007: 627-631.
8. Efron N. Clinical application of grading scales for contact lens complications. *Optician*. 1997; 213:26-35.
9. Pérez-Cabré E, Millán MS, Abril HC, Valencia E. Image analysis of contact lens grading scales for objective grade assignment of ocular complications. *Proc of SPIE*. 2005; 5287:418-427.
10. Efron N, Morgan PB, Jagpal R. The combined influence of knowledge, training and experience when grading contact lens complications. *Ophthal Physiol Opt*. 2003; 23:79-85.
11. Efron N, Morgan PB, Farmer C, Furuborg J, Struk R, Carney LG. Experience and training as determinants of grading reliability when assessing the severity of contact lens complications. *Ophthal Physiol Opt*. 2003; 23:119-124.
12. Bailey IL, Bullimore MA, Raasch TW, Taylor HR. Clinical grading and the effects of scaling. *Invest Ophthalmol Vis Sci*. 1991; 32:422-432.
13. Efron N, Morgan PG, Katsara SS. Validation of grading scales for contact lens complications. *Ophthal Physiol Opt*. 2001; 21:17-29.
14. Welkowitz J, Ewen RB, Cohen J. *Introductory Statistics for Behavioural Sciences*, 3rd ed. New York, USA: Academic Press; 1982: 251.

15. ISO, 5725-1 Accuracy (trueness and precision) of measurement methods and results – part 1: general principles and definitions. International Organization for Standardization, Switzerland.
16. Scott J, Huskisson EC. Graphic representation of pain. *Pain*. 1976; 2:75-84.
17. Efron N, McCubbin S. Grading contact lens complications under time constraints. *Optom Vis Sci*. 2007; 84:1082-1086.

FIGURES

Figure 1: Example of a Vertical Visual Analogue Scale (VVAS) used for the evaluation of bulbar hyperaemia

Figure 2:

Figure 2 a: Grading bulbar hyperaemia. Image under evaluation (right) and end of scale clues (left)

Figure 2 b: Grading limbal vascularisation. Image under evaluation (right) and end of scale clues (left)

Figure 2 c: Grading lid redness. Image under evaluation (right) and end of scale clues (left)

Figure 3:

Figure 3 a: Overall grading performance for each knowledge group (Means and 95% LSD intervals)

Figure 3 b: Overall grading scores for each condition (Means and 95% LSD intervals)

Figure 4:

Figure 4 a: Interaction plot for grading scores and condition, depicting grading behaviour

Figure 4 b: Residual plot of the grading scores for each knowledge group

Declaration of Interest

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.